ARCHIVES OF ELECTRICAL ENGINEERING VOL. 75(1), pp. 117–134 (2026)

DOI 10.24425/aee.2026.156805

Application of spectrogram fusion and deep learning for hand gesture classification based on forearm electromyographic signals

JAN RAK, PIOTR GAS ✉

Department of Electrical and Power Engineering
Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering
AGH University of Krakow
30 Mickiewicza Ave., 30-059 Krakow, Poland
e-mail: janekraktu@student.agh.edu.pl, ✉ *piotr.gas@agh.edu.pl*

(Received: 11.07.2025, revised: 27.01.2026)

Abstract: Hand gesture recognition based on surface electromyographic (sEMG) signals plays a critical role in modern human–computer interaction systems, particularly in upper-limb prosthetic applications. This study presents a method for classifying six selected hand gestures using sEMG signals acquired from three forearm muscles. The recorded signals were digitally filtered, and an automatic segmentation algorithm was developed to isolate individual gestures from the continuous muscle activity recordings. These segments were transformed into spectrograms using the short-time Fourier transform (STFT), which served as input data for various convolutional neural network (CNN) architectures. The study compares two approaches to data processing: one in which signals from each channel were analyzed separately, and another in which spectrograms from all three channels were fused into a single three-channel input. The primary objective was to investigate which method better captures the inter-relationships between the activity patterns of different muscles. The models were trained and evaluated using cross-validation. The best-performing architecture achieved an accuracy of 99%. The results indicate that fusing spectrograms from multiple channels into a single input can enhance the classification performance of complex muscle activity patterns, particularly when the amount of available training data is limited.

Key words: classification, convolutional neural network (CNN), electromyographic (EMG) signals, hand gesture recognition, spectrogram fusion

1. Introduction

The classification of electromyographic (EMG) signals has become a central focus in biomedical and engineering studies, particularly in applications involving prosthetic control and human–computer interfaces (HCI) [1]. EMG signals contain valuable information about neuromuscular activity, making them a widely used tool for analyzing muscle function in both clinical and engineering applications [2] and making them well-suited for recognizing hand gestures that can be used to control assistive technologies. Beyond prosthetic and HCI applications, hand gesture recognition also plays a critical role in sign language interpretation systems [3] and other bio-signal interfaces aimed at restoring or enhancing motor function [4]. People with upper limb amputations face numerous challenges in daily life, including the loss of fine motor control and independence in daily tasks. Advanced prosthetic devices capable of interpreting EMG signals offer the potential to restore lost functionality by accurately detecting user intentions through gesture recognition. However, current commercially available prostheses often suffer from several limitations such as mechanical complexity, insufficient robustness, signal noise sensitivity, and difficulty in achieving intuitive control. These challenges are further compounded by variability in EMG signals between users, differences in skin impedance, and difficulties with consistent sensor placement. To address these issues, robust and adaptive signal acquisition and classification techniques are required. A key consideration in EMG-based gesture recognition is the selection of appropriate muscle groups for signal acquisition. The forearm is frequently targeted due to its high concentration of muscles involved in wrist and finger movements. Even in cases of transradial amputation, viable EMG signals can often be recorded from residual muscles in the stump, allowing control interfaces to be implemented even after limb loss [5]. Over the past few decades, various methods have been developed to tackle the EMG signal classification problem. Traditional approaches rely on manually engineered features derived from the time, frequency, or time-frequency domains. Common time-domain features include mean absolute value (MAV), root mean square (RMS), variance, and waveform length. In the frequency domain, features are often computed using the fast Fourier transform (FFT), short-time Fourier transform (STFT), or wavelet transforms, allowing the extraction of metrics such as median frequency (MDF), mean frequency (MNF), and power spectral density (PSD) [6]. Early signal analysis approaches relied heavily on these handcrafted features to classify muscle activity. Reaz *et al.* [7] provided a comprehensive survey of such traditional methods, highlighting their computational simplicity but also their limited ability to generalize across users or adapt to the nonlinear and nonstationary nature of EMG signals. Although these handcrafted features are computationally efficient and suitable for real-time systems, they often exhibit poor generalization across users or sessions and a limited ability to capture nonlinear patterns. Moreover, the assumption of stationarity required by many spectral methods does not always hold for EMG, which is inherently nonstationary and stochastic in nature. To improve recognition stability and robustness, filtering techniques such as variational mode decomposition have been introduced. Ma *et al.* [8] demonstrated that effective signal denoising prior to classification could significantly enhance accuracy, particularly in the presence of noise or during dynamic movement conditions. In response to these limitations, deep learning methods, widely applied across various areas of bioengineering [9–13], have gained significant traction in the field of gesture recognition [14]. These approaches automatically extract hierarchical and abstract representations (referred to as deep features) directly from raw or minimally processed

data. Convolutional neural networks (CNNs), in particular, have demonstrated strong performance in EMG-based gesture recognition tasks. By converting EMG signals into 2D spectrograms using STFT, CNNs are able to effectively model spatial and temporal signal patterns, bypassing the need for manual feature engineering [15]. Several studies have demonstrated the effectiveness of CNNs and other deep learning techniques for this task. CNNs were used to recognize basic hand gestures, and their superiority over classical machine learning approaches was confirmed in [16]. Further improvements were shown by [17], where CNNs with varied input dimensions achieved highly accurate results. Yamanoi *et al.* [5] showed that CNNs can outperform conventional classifiers in distinguishing hand postures. Cote-Allard *et al.* [18] extended this work by applying transfer learning strategies to reduce model training time and improve robustness across users. More advanced architectures have also been proposed to improve practical usability. Zhai *et al.* [19] introduced a self-adaptive model capable of recalibrating without the need for manual intervention, improving long-term stability. To further capture temporal dependencies in EMG signals, Sun *et al*. [20] employed dilated LSTM networks to model the dynamics of gesture transitions. Hybrid deep learning frameworks that combine spatial and temporal modeling have also emerged; Hu *et al*. Paper [21] proposed an attention-enhanced CNN-RNN architecture, improving both classification performance and model interpretability in sequential EMG data. The role of hyperparameter tuning in building robust and stable classification models was explored in the study [22]. Recurrent architectures have also gained attention. For example, investigation [23] proposed using sequences of EMG windows for gesture prediction. Hybrid models, which combine handcrafted features and deep learning, have also shown promise. One such approach in [24] merged both types of features into a single classifier. Additionally, study [25] explored integrating CNNs with long short-term memory (LSTM) layers to incorporate temporal dependencies into the model's decision-making process. Recent studies have further advanced these architectures by focusing on real-time deployment, robustness, and application-specific adaptability. For instance, CNN-based models trained on EMG-derived frequency features have shown strong performance in guiding robotic arms with high precision [26]. Other approaches have leveraged feed-forward neural networks for real-time EMG pattern recognition in embedded systems [27]. In addition, fully embedded systems leveraging high-density sEMG data and deep learning classifiers have demonstrated adaptive, low-latency performance suitable for wearable control interfaces [28]. Similar efforts to deploy robust CNN models on multicore IoT platforms have enabled efficient gesture recognition in power-constrained environments [29]. Comparative analyses with traditional artificial neural network (ANN)-based systems further confirm the superiority of deep learning approaches under variable signal conditions [30]. Moreover, multi-sensor systems and ensemble strategies have been explored to improve classification reliability using minimal EMG channels [31]. Collectively, these contributions demonstrate a clear trend toward integrating adaptive, deep learning-based EMG interfaces into real-time assistive technologies. Efficient implementation of such systems on edge devices, has also become an important focus in recent studies [32–37]. A variety of machine learning algorithms has also been used in other important applications [38–43]. Previous studies, such as Geng *et al*. [44] and Tepe *et al*. [45], used similar single-subject setups and achieved high accuracy with CNN and support vector machine (SVM) models, respectively. However, these works did not explore how combining spectrograms from multiple EMG channels could improve feature representation or capture inter-muscle relationships. In this study, we compare early and late spectrogram fusion strategies to evaluate their impact on classification performance.

In the current paper, we present a complete pipeline for hand gesture classification based on EMG signals and deep learning techniques. EMG data were collected from three superficial forearm muscles while subjects performed six distinct hand gestures. A custom segmentation algorithm based on amplitude thresholding was applied to extract active gesture segments from continuous recordings. Each segment was transformed into a spectrogram using the STFT, yielding a three-channel $3 \times 129 \times 64$ tensor for each instance. To evaluate model performance, five CNN architectures were implemented. The first four models used a single-branch structure that processes the entire three-channel spectrogram as a unified input. This early fusion strategy enables the network to convolve over all channels simultaneously, similar to processing an RGB image, and allows the model to learn joint spatial and temporal patterns across multiple muscles. By capturing inter-muscle coordination and shared activation patterns, this approach supports the extraction of richer, more informative features, which can be especially useful for distinguishing complex or subtle gestures. Additionally, a fifth architecture was introduced using a three-branch CNN model that processes each spectrogram channel independently before merging the extracted features. This late fusion strategy allows the network to specialize in channel-specific patterns while still learning complementary information during the fusion stage. Such modular representation can be advantageous when signal variability across channels is significant or when muscle-specific distinctions are critical to accurate gesture classification. The results highlight the potential of early and late fusion strategies in deep learning models for EMG-based gesture recognition and support their integration into responsive, intelligent human–machine interaction systems.

2. Materials and methods

Electrical signals generated by muscle contractions are characterized by their low amplitude and nonrepeatability. Surface electromyography (sEMG) measurements can vary significantly between individuals due to differences in skin condition, preparation of the measurement site, and electrode placement. Additionally, surface EMG is highly dependent on individual anatomical features, making consistent electrode placement across subjects challenging. Achieving accurate and repeatable measurements often requires a personalized approach tailored to each subject. Readings from surface electrodes represent a composite EMG signal, reflecting the activity of multiple muscles located beneat the skin at the electrode site. As a result, the recorded signal may include noise or irrelevant information from surrounding muscles, which can affect the accuracy of the data. To mitigate these effects, the gestures selected for classification were chosen to involve distinct muscle activation patterns and to be performed by superficial muscles. The selected gestures are illustrated in Fig. 1.

All EMG signals were acquired using the Biomonitor ME6000. In every recording signals from three channels were captured, each channel consisting of two measuring electrodes and one reference electrode. Every channel was used to acquire signals from one of three muscles: *flexor carpi radialis*, *flexor carpi ulnaris* and *extensor digitorum communis*. Silver/silver chloride (Ag/AgCl) electrodes were adhered to shaved and cleaned skin above the chosen muscles. The planned electrode placement is illustrated in Fig. 2, while a close-up photograph showing the electrodes adhered and wired during one of the recording sessions is presented in Fig. 3. In support of electrode placement and skin preparation protocols, we follow established guidelines [46], ensuring consistent capture from the flexor and extensor forearm muscles.

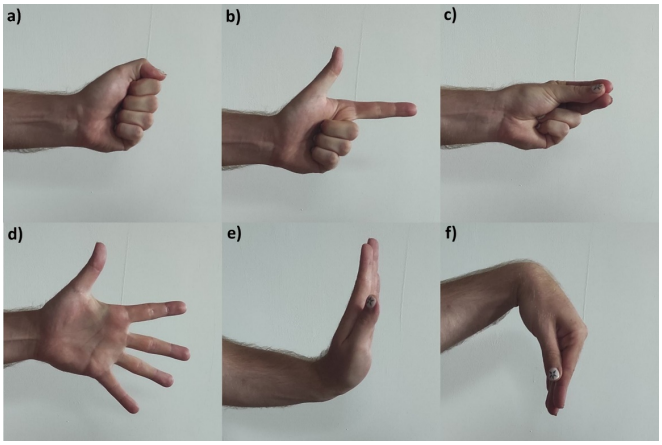

Fig. 1. Gestures chosen for classification: (a) finger flexion; (b) pointing; (c) three-finger pinch; (d) finger abduction; (e) wrist extension; (f) wrist flexion

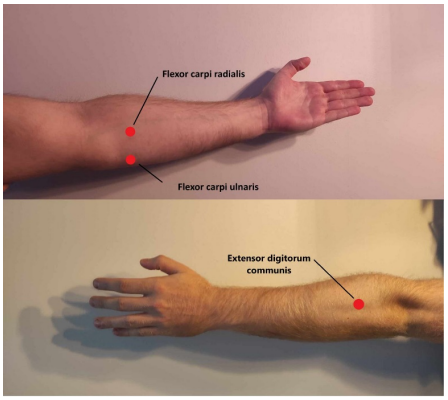

Fig. 2. Planned electrode placement based on the three chosen muscles responsible for hand movements

(a) (b)

Fig. 3. Measurements of hand EMG signals: (a) Biomonitor ME6000; (b) wired measuring channels placed at selected muscles (*flexor carpi radialis, flexor carpi ulnaris*)

The signals were captured with a sampling frequency of 1 000 Hz. Signal acquisition took the form of recordings in which the participant performed 20 repetitions of a chosen gesture. Each recording started with a 2-second-long period of no muscular activity. After that, the participant performed 2-second-long gestures alternately with 3-second-long resting periods. Collected recordings for each gesture corresponded to 140 repetitions of that gesture, which altogether summed up to three-channel EMG signals of 840 gesture repetitions. Recorded signals were segmented using a sliding window algorithm. All recordings were collected with the cooperation of a single participant. This approach was intentional and aligned with the single-subject study design.

The presented algorithm windows all three channels of a single signal recording simultaneously with an overlap of $O = 75$ samples using a sliding window with a fixed length of $L = 350$ samples. As emphasized in paper [47], segmenting nonstationary EMG into time windows reduces stochastic noise and improves classification robustness.

Let $x_c[n]$ denote the raw EMG signal from channel $c \in \{1, 2, 3\}$. Each signal was first rectified to obtain the discrete time series representing the rectified signal:

$$r_c[n] = |x_c[n]| . \tag{1}$$

For every window

$$w_i = \{r_c[n_i], r_c[n_i + 1], \ldots, r_c[n_i + L - 1]\} , \tag{2}$$

the mean rectified amplitude (MRA) was computed as:

$$A_{c,i} = \frac{1}{L} \sum_{k=0}^{L-1} r_c[n_i + 1], \tag{3}$$

where L is the window length, n_i is the index of the first sample of the window.

A baseline amplitude $\bar{A}_{c,0}$ was estimated from the first $N_0 = 1\,000$ rectified samples of each channel as:

$$A_{c,0} = \frac{1}{N_0} \sum_{k=0}^{N_0-1} r_c[k]. \tag{4}$$

The activation threshold for each channel was defined as:

$$T_c = 1.3 A_{c,0}. \tag{5}$$

Gesture onset and offset points were determined by threshold crossings of the mean rectified amplitude. A gesture onset was detected when the amplitude in any of the three channels exceeded its corresponding threshold:

$$A_{c,i-1} < T_c \text{ and } A_{c,i} \geq T_c, \tag{6}$$

and a gesture offset was detected when the amplitudes from all of the three channels fell below their corresponding thresholds:

$$A_{c,i-1} \geq T_c \text{ and } A_{c,i} < T_c. \tag{7}$$

The index corresponding to the start of window w_i was marked as the onset or offset point, respectively. To ensure reliable segmentation and prevent false detections due to transient fluctuations, a minimum separation constraint of $D_{\min} = 250$ samples was imposed between

gesture onset and offset, as well as between consecutive gesture segments. Using criteria presented by above Eqs. (1)–(7), gesture segments were reliably extracted from the original EMG signals. The automatic segmentation algorithm, which uses a 75-sample overlap and threshold-based amplitude detection, follows methodologies outlined in studies [48] and [49]. These sources demonstrate how overlapping sliding windows and envelope-based thresholds effectively extract gesture segments from continuous EMG data. The segmented portions of acquired data were then combined to form the dataset used for further analysis and classification. An example result of automatic segmentation can be seen in Fig. 4.

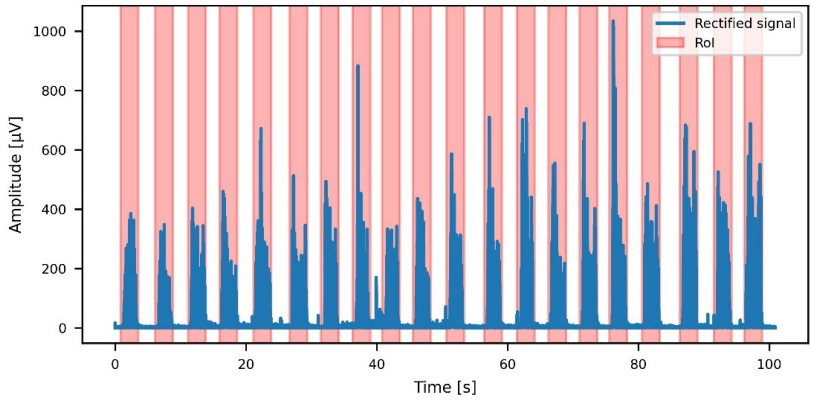

Fig. 4. Plot representing the results of applying the segmentation algorithm on the rectified signal recording. The plot represents an example segmentation of the signal recorded from the *extensor digitorum communis* while performing wrist extension

2.1. Signal pre-processing

For signal pre-processing, three main procedures were carried out. Firstly, the obtained segmented signals were cropped and then padded to fit the same length of 4 000 samples. The segmented EMG signals vary in length because it is difficult for individuals to perform gestures that last exactly two seconds. Consistent signal length is crucial for efficient data handling and machine learning training and validation procedures. Cropping was chosen for this task so as not to change the spectral features of the signals, which may occur while using signal resampling. After the cropping procedure, all of the signals were filtered. The filtering was performed using a 5th-order Butterworth digital bandpass filter with cutoff frequencies of 20 Hz and 400 Hz for the lower and upper limits, respectively [50]. The filtered and cropped signals were used to generate spectrograms using the STFT. This time-frequency analysis technique allows for the examination of how the frequency content of a signal changes over time. It operates by dividing the signal into overlapping segments and applying the Fourier transform to each segment independently. This decomposition enables localized frequency analysis across the time axis, making STFT particularly useful for non-stationary signals such as EMG. In the STFT process, a sliding window function is applied to isolate short portions of the signal. Each windowed segment is then transformed into the frequency domain using the discrete Fourier transform (DFT). The result is a two-dimensional representation known as a spectrogram, where one axis represents time, the other refers to the frequency, and the color or intensity represents signal amplitude at each time-frequency coordinate.

Mathematically, the STFT of a signal $x(t)$ is defined as [51]:

$$\mathrm{STFT}_x(\tau y) = \int_{-\infty}^{+\infty} x(t)\gamma(t-\tau)e^{-j2\pi ft}\mathrm{d}t, \tag{8}$$

where $\gamma(t-\tau)$ is the window function centered at time τ, and f is the frequency.

To compute the STFT in practice, discrete parameters are selected. In this study, a sampling frequency of 1 kHz was used, with each signal segment consisting of 129 samples and an overlap of 68 samples between segments. The FFT length was set to 256 points. These parameters were chosen to balance frequency resolution with computational efficiency. After transformation, each EMG signal is represented as a 3D matrix with dimensions $3 \times 129 \times 64$ corresponding to the number of channels and the size of the spectrogram. An exemplary input of three spectrograms, with amplitude presented using color mapping, can be seen in Fig. 5.

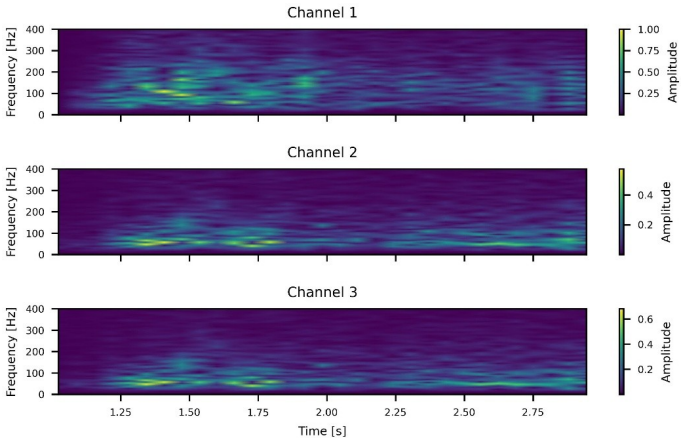

Fig. 5. Spectrograms generated from signals acquired from one-gesture repetitions

The spectrograms shown above are particularly suitable for implementing into CNNs, as they effectively convert the EMG signal into an image-like format, enhancing the model's ability to learn spatial and frequency-related features for classification tasks.

2.2. Models and training

In recent literature, convolutional neural networks (CNNs) have become the most prevalent deep learning approach for EMG signal classification, often utilizing image-based representations of time-frequency features. Following this trend, the present work applies a CNN-based architecture for classifying EMG spectrograms generated using the short-time Fourier transform (STFT). Each input instance is represented as a $3 \times 129 \times 64$ matrix, corresponding to spectrograms derived from three distinct EMG channels located on the forearm. To facilitate effective training and generalization, 5-fold cross-validation was employed on the dataset, which comprised 769 manually validated EMG recordings. This approach ensures that each sample is used for both training

and evaluation, allowing for a robust and independent assessment of the model's performance on unseen data. The composition of the collected dataset can be seen in Fig. 6. The dataset is well-balanced across gesture classes, which is essential for ensuring unbiased training [52]. The proposed single-branch CNN models vary in architectural complexity, consisting of one to four convolutional blocks, each composed of a convolutional layer, a normalization layer, and a ReLU activation. As the number of blocks increases, so does the model's capacity to extract higher-order spatial and temporal features from the spectrograms. A global average pooling layer and a fully connected output layer finalize the classification process across six predefined gesture classes.

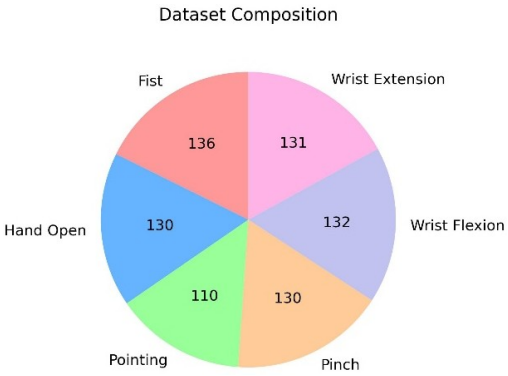

Fig. 6. Composition of the obtained dataset

Figure 7 represents the structure of the most complex model of the four implemented single-branch models. The presented model consists of 98 694 trainable parameters. The single-branch models take the full three-channel spectrogram tensor as input and perform convolution over the entire tensor, effectively fusing information from spectrograms of signals recorded from different muscles, similar to how a three-channel RGB image is processed in standard CNNs. This early fusion strategy allows the model to jointly learn spatial and temporal patterns across all muscle signals, potentially capturing inter-muscle correlations and coordinated activation patterns that might be overlooked in separate processing [53]. As a result, the model can exploit richer feature representations, which can improve classification performance, especially in tasks involving complex or subtle gesture distinctions. Additionally, a three-branch architecture was implemented to investigate whether processing each spectrogram channel independently, prior to fusion, could yield improved classification performance.

In this setup, each branch is dedicated to one muscle signal and extracts features from its corresponding spectrogram, independently, allowing the model to specialize in capturing channel-specific patterns. The extracted features are then combined at a later stage, enabling the model to learn both individual and complementary information across channels. This late fusion approach provides a more modular representation of muscle activity and may be particularly advantageous when inter-channel variability is high or when subtle distinctions in signal characteristics are critical for accurate gesture recognition.

Figure 8 presents the proposed three-branch model structure consisting of 295 206 trainable parameters. During training, a cross-entropy loss function was used to compare predicted class distributions to ground truth labels. Optimization was performed using the Adam algorithm with

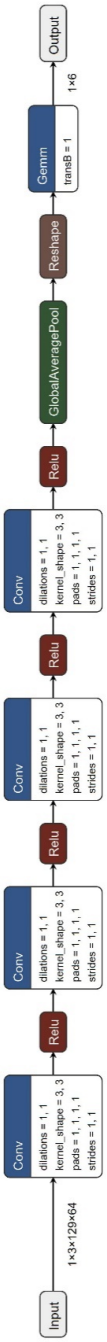

Fig. 7. Structure of the most complex single-branch model

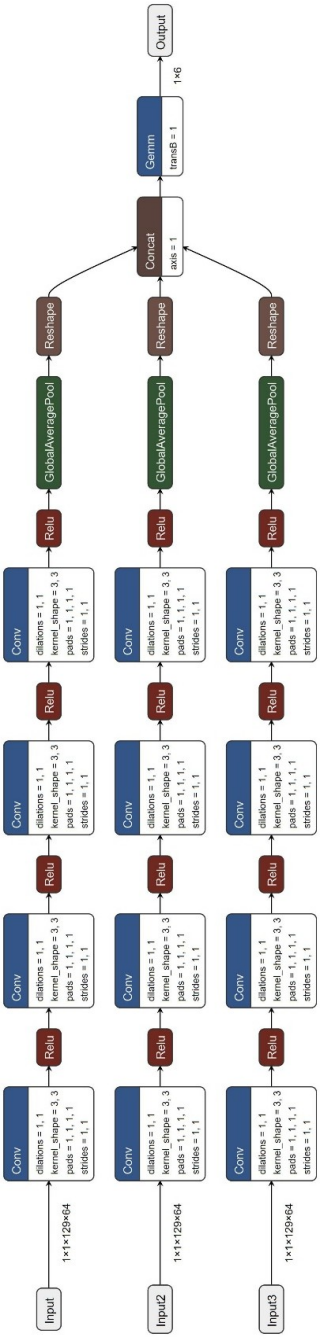

Fig. 8. Structure of the three-branch model

a learning rate of 0.003 over 50 epochs for each fold. All spectrogram inputs were normalized using min-max scaling to improve convergence stability and overall training efficiency [54]. Each model's learning dynamics were monitored through loss function curves across epochs, and the results demonstrated effective convergence without signs of overfitting. These trends were consistent for both training and validation datasets, suggesting robust generalization capabilities. All the CNN models were implemented and trained using the PyTorch library [55].

3. Results and discussion

As demonstrated in the studies by Naik *et al*. [56], Gandolla *et al*. [57], as well as Yaman and Subasi [58], evaluation metrics such as accuracy, recall, specificity, precision, negative predictive value (NPV), area under the receiver operating characteristic curve (AUROC), and F_1-score are effective for assessing the performance of an EMG classifier. To evaluate the performance of the trained CNN models, classification metrics were computed on both the training and test datasets. While the training set performance was assessed primarily through the generation of confusion matrices, the test set was used for a more comprehensive evaluation. All results were generated using Python, with the Seaborn library utilized to create the confusion matrices for visual inspection of prediction accuracy per class. Model evaluation on the test dataset was conducted using the four performance metrics: precision, recall (sensitivity), F_1-score, and accuracy. These metrics provide insight not only into the overall performance of the classifier but also into its behavior across individual gesture classes. The definitions of the metrics are as follows [20, 46]:

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{9}$$

$$\text{Recall(Sensitivity)} = \frac{TP}{TP + FN}, \tag{10}$$

$$F_1 = 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{11}$$

$$\text{Accurancy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{12}$$

where TP, FP, FN, and TN refer to the true positives, false positives, false negatives, and true negatives, respectively.

Parameters, defined by above Eqs. (9)–(12), are derived from the confusion matrix, which compares actual labels to predicted labels for each class. Precision reflects the proportion of correctly predicted samples among all predictions for a given class. Recall indicates the model's ability to correctly identify all instances of a particular class. The F_1-score balances precision and recall, serving as a robust metric especially in imbalanced datasets. Accuracy provides a general measure of the percentage of correctly classified samples across the entire test set.

Figure 9 presents the confusion matrices evaluating the performance of implemented models including spectrogram fusion model with one, two, three and four convolution layers as well as the three-branch model. Detailed results are summarized in Table 1. The presented results demonstrate a clear and consistent progression in classification performance, with accuracy steadily improving from Model 1 (81%) to Model 4 (99%). This upward trend reflects the benefits of

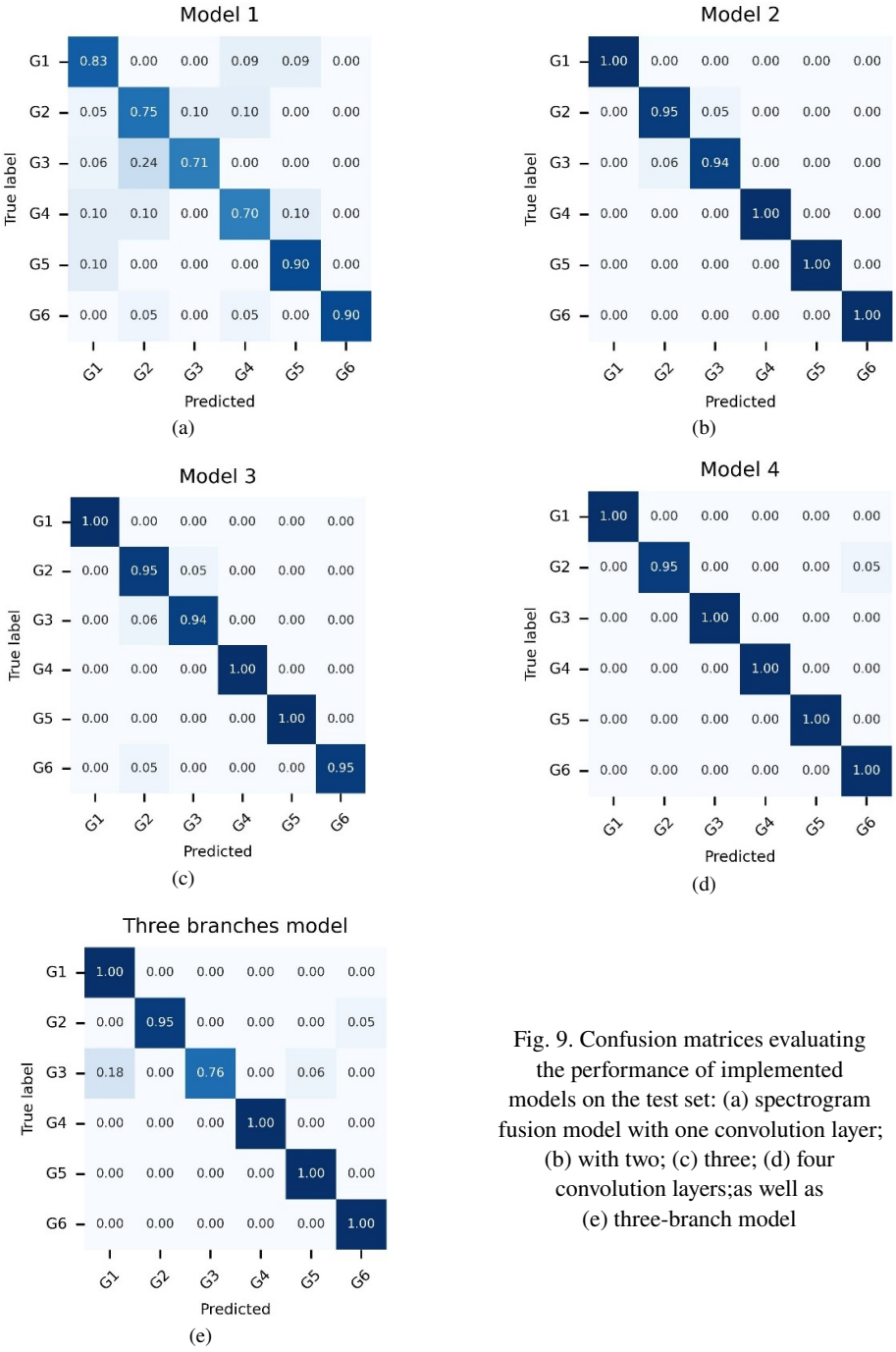

Fig. 9. Confusion matrices evaluating the performance of implemented models on the test set: (a) spectrogram fusion model with one convolution layer; (b) with two; (c) three; (d) four convolution layers;as well as (e) three-branch model

increasing model complexity and refining feature extraction strategies. As deeper architectures and more sophisticated processing were introduced, the models became better at capturing subtle patterns in EMG signals, resulting in higher accuracy and more consistent class-wise performance. Three-branch model results did not stand out much from early fusion models and were weaker in comparison to smaller models using the early fusion. While the three-branch model performed reliably, it did not outperform the early fusion models and was slightly weaker than some of the simpler early fusion models. This indicates that its added complexity did not lead to better results for this task. To place these findings in context, Table 2 compares the performance of our models with results from related studies involving similar subject groups. The confusion matrices provided qualitative insights into model behavior, revealing which gestures were most reliably recognized and where misclassifications tended to occur. In particular, misclassifications between similar or adjacent gestures often suggest overlapping representations in the learned feature space, especially in earlier, less complex models. While overall performance improved significantly with each architectural refinement, gesture-specific challenges remained evident. Wrist extension was consistently the most accurately classified gesture, while more complex or subtle gestures, such as hand open, pinch, and pointing were more prone to confusion, particularly in simpler models. These discrepancies can be explained by considering the anatomical and physiological characteristics of the underlying musculature.

Table 1. Summary of evaluation results for all analyzed models

Analyzed model	**Average precision**	**Average recall**	**Average F_1-score**	**Accuracy**
Model 1	0.82	0.81	0.81	0.81
Model 2	0.98	0.98	0.98	0.98
Model 3	0.97	0.97	0.97	0.97
Model 4	0.99	0.99	0.99	0.99
Three-branch model	0.96	0.95	0.95	0.96

Table 2. Comparison the results from related studies involving similar subject groups with current research

Study	**Subjects**	**Number of gestures**	**Channels**	**Classifier**	**Accuracy**
Benalcazar *et al.* (2018) [22]	1	5	2	ANN	0.978
Zhang *et al.* (2019) [30]	1	6	2	ANN	0.987
Geng *et al.* (2016) [44]	1	8	N/A	CNN	0.990
Tepe *et al.* (2022) [45]	1	5–7	8	C-SVM	0.991
Current work	1	6	3	CNN	0.990

EMG signals were collected from two muscles primarily responsible for wrist movements and only one associated with finger movements. Wrist gestures engage larger, more superficial muscles that generate stronger and more distinct EMG signals, which are easier to detect and classify. In

contrast, finger gestures depend on smaller, deeper, and more complex muscle groups that produce weaker and more variable signals, which are more likely to overlap and result in classification errors. This disparity in signal quality and separability accounts for the consistent advantage seen in wrist gesture recognition. Overall, the findings confirm that with appropriate feature engineering and parameter tuning, even classical classifiers using handcrafted features can achieve high levels of accuracy comparable to deep learning models when applied to EMG-based gesture recognition involving a limited set of six gestures. This is particularly advantageous in applications where computational efficiency, simplicity, or limited training data are critical constraints. However, several limitations must be acknowledged. The dataset lacked participant diversity, reducing confidence in generalizability across users. The system was evaluated offline, without real-time testing, and its performance under varying conditions, such as electrode displacement, inter-session variability, or signal noise, was not systematically analyzed. Furthermore, the gesture vocabulary was limited, which may constrain the applicability of the system in more complex, real-world scenarios. Future work should aim to address these issues by testing the system in dynamic, multi-user environments, incorporating adaptive learning strategies, and expanding the gesture set to better support practical deployment. Additionally, EMG channels focused on finger-specific muscles may also help improve performance for more intricate gestures, enhancing the robustness and versatility of EMG-based interfaces.

4. Conclusions

This study presented a complete deep learning-based pipeline for classifying hand gestures using surface electromyographic (EMG) signals recorded from three superficial forearm muscles. By converting segmented EMG data into short-time Fourier transform (STFT) spectrograms and leveraging convolutional neural networks (CNNs), we demonstrated the feasibility and effectiveness of automatic gesture recognition without the need for handcrafted features. The proposed models achieved high classification accuracy across six predefined gestures, confirming the strength of CNNs in capturing complex spatiotemporal patterns inherent in EMG signals. Among the five tested CNN architectures, deeper models with more convolutional blocks exhibited superior performance, with the most complex model reaching 99% accuracy on the test set. However, Model 3 offered a compelling trade-off between classification performance and architectural complexity, making it suitable for practical deployment in resource-constrained environments. Furthermore, the comparison between early fusion (single-branch) and late fusion (multi-branch) strategies underscored the benefits of learning both joint and channel-specific features, especially when muscle coordination and inter-channel variability play significant roles in gesture discrimination. Despite these promising results, the system's generalization capability requires further validation through testing with a larger and more diverse user group, including individuals with limb loss. Moreover, real-time integration remains an important direction for future development. Expanding the number of recognizable gestures, incorporating adaptive learning techniques to address inter-user variability, and exploring alternative network architectures, such as CNN-RNN hybrids may further improve system responsiveness and robustness. Ultimately, this work contributes toward the advancement of intuitive, non-invasive control systems for prosthetics and human-machine interfaces, leveraging the power of deep learning to interpret EMG signals with high precision and adaptability.

References

[1] Buchwald M., Jukiewicz M., *Project and evaluation EMG/EOG human-computer interface*, Przeglad Elektrotechniczny, vol. 93, no. 7, pp. 128–131 (2017), DOI: 10.15199/48.2017.07.28.

[2] Bodera P., Stankiewicz W., Kalicki B., Kieliszek J., Sobiech J., Krawczyk A., *The surface electromyography biofeedback in pain management-theoretical assumptions and possibilities of using the method*, Przeglad Elektrotechniczny, vol. 88, no. 12B, pp. 115–116 (2012); available online at: http://pe.org.pl/articles/2012/12b/34.pdf.

[3] Rastogi U., Mahapatra R.P., Kumar S., *Advancements in machine learning techniques for hand gesture-based sign language recognition: A comprehensive review*, Archives of Computational Methods in Engineering (2025), DOI: 10.1007/s11831-025-10258-z.

[4] Kawala-Sterniuk A., Browarska N., Al-Bakri A., Pelc M., Zygarlicki J., Sidikova M., Martinek R., Gorzelanczyk E.J., *Summary of over Fifty Years with Brain-Computer Interfaces–A Review*, Brain Sciences, vol. 11, iss. 1, no. 43 (2021), DOI: 10.3390/brainsci11010043.

[5] Yamanoi Y., Ogiri Y., Kato R., *EMG-based posture classification using a convolutional neural network for a myoelectric hand*, Biomedical Signal Processing and Control, vol. 55, no. 101574 (2020), DOI: 10.1016/j.bspc.2019.101574.

[6] Tsai A.C., Luh J.J., Lin T.T., *A novel STFT-ranking feature of multi-channel EMG for motion pattern recognition*, Expert Systems with Applications, vol. 42, no. 7, pp. 3327–3341 (2015), DOI: 10.1016/j.eswa.2014.11.044.

[7] Raez M.B.I., Hussain M.S., Mohd-Yasin F., *Techniques of EMG signal analysis: detection, processing, classification and applications*, Biological Procedures Online, vol. 8, pp. 11–35 (2006), DOI: 10.1251/bpo115.

[8] Ma S., Lv B., Lin C., Sheng X., Zhu X., *EMG Signal Filtering Based on Variational Mode Decomposition and Sub-Band Thresholding*, IEEE journal of biomedical and health informatics, vol. 25, no. 1, pp. 47–58 (2020), DOI: 10.1109/JBHI.2020.2987528.

[9] Budzynska J., Kujawa M., Roszczyk R., *Applying artificial intelligence techniques in computed tomography for supporting liver cancer diagnosis*, 2024 Progress in Applied Electrical Engineering (PAEE), IEEE Xplore, pp. 1–6 (2024), DOI: 10.1109/PAEE63906.2024.10701445.

[10] Altintas G., Yasar H., Uslu I.E., Demirel Y., Joof S., Akinci M.N., Yilmaz T., Akduman I., *Antenna Array Optimization via Deep Learning For Breast Cancer Microwave Hyperthermia Application: Preliminary Results*, 2022 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting (AP-S/URSI), IEEE Xplore, pp. 697–698 (2022), DOI: 10.1109/AP-S/USNC-URSI47032.2022.9887012.

[11] Glowacz A., Glowacz Z., *Recognition of images of finger skin with application of histogram, image filtration and K-NN classifier*, Biocybernetics and Biomedical Engineering, vol. 36, no. 1, pp. 95–101 (2016), DOI: 10.1016/j.bbe.2015.12.005.

[12] Kawa B., Borkowski P., Rodak M., *Building management system based on brain computer interface. Review*, Archives of Electrical Engineering, vol. 70, no. 4, pp. 887–905 (2021), DOI: 10.24425/aee.2021.138268.

[13] Azad M.M., Raouf I., Sohail M., Kim H.S., *Structural Health Monitoring of Laminated Composites Using Lightweight Transfer Learning*, Machines, vol. 12, iss. 9, no. 589 (2024), DOI: 10.3390/machines12090589.

[14] Sarowar M.S., Farjana N.E.J., Khan M.A.I., Mutalib M.A., Islam S., Islam M., *Hand Gesture Recognition Systems: A Review of Methods, Datasets, and Emerging Trends*, International Journal of Computer Applications, vol. 197, no. 2, pp. 1–33 (2025), DOI: 10.5120/ijca2025924776.

[15] Atzori M., Cognolato M., Muller H., *Deep Learning with Convolutional Neural Networks Applied to Electromyography Data: A Resource for the Classification of Movements for Prosthetic Hands*, Frontiers in Neurorobotics, vol. 10, no. 9 (2016), DOI: 10.3389/fnbot.2016.00009.

[16] Ding Z., Yang C., Tian Z., Yi C., Fu Y., Jiang F., *sEMG-Based Gesture Recognition with Convolution Neural Networks*, Sustainability, vol. 10, iss. 6, no. 1865 (2018), DOI: 10.3390/su10061865.

[17] Asif A.R., Waris A., Gilani S.O., Jamil M., Ashraf H., Shafique M., Niazi I.K., *Performance Evaluation of Convolutional Neural Network for Hand Gesture Recognition Using EMG*, Sensors, vol. 20, iss. 6, no. 1642 (2020), DOI: 10.3390/s20061642.

[18] Cote-Allard U., Fall C.L., Drouin A., Campeau-Lecours A., Gosselin C., Glette K., Laviolette F., Gosselin B., *Deep learning for electromyographic hand gesture signal classification using transfer learning*, IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 27, no. 4, pp. 760–771 (2019), DOI: 10.1109/TNSRE.2019.2896269.

[19] Zhai X., Jelfs B., Chan R.H.M., Tin C., *Self-recalibrating surface EMG pattern recognition for neuroprosthesis control based on convolutional neural network*, Frontiers in Neuroscience, vol. 11, art. no. 379 (2017), DOI: 10.3389/fnins.2017.00379.

[20] Sun T., Hu Q., Libby J., Atashzar S.F., *Deep Heterogeneous Dilation of LSTM for Transient-Phase Gesture Prediction Through High-Density Electromyography*, IEEE Robotics and Automation Letters, vol. 7, no. 2. pp. 2851–2858 (2022), DOI: 10.1109/LRA.2022.3142721.

[21] Hu Y., Wong Y., Wei W., Du Y., Kankanhalli M., Geng W., *A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition*, PLoS One, vol. 13, iss. 10, no. e0206049 (2018), DOI: 10.1371/journal.pone.0206049.

[22] Barona Lopez L.I., Ferri F.M., Zea J., Valdivieso Caraguay A.L., Benalcazar M.E., *CNN-LSTM and post-processing for EMG-based hand gesture recognition*, Intelligent Systems with Applications, vol. 22, no. 200352 (2024), DOI: 10.1016/j.iswa.2024.200352.

[23] Zhang Z., He C., Yang K., *A Novel Surface Electromyographic Signal-Based Hand Gesture Prediction Using a Recurrent Neural Network*, Sensors, vol. 20, iss. 14, no. 3994 (2020), DOI: 10.3390/s20143994.

[24] Fajardo J.M., Gomez O., Prieto F., *EMG hand gesture classification using handcrafted and deep features*, Biomedical Signal Processing and Control, vol. 63, no. 102210 (2021), DOI: 10.1016/j.bspc.2020.102210.

[25] Chen J., Bi S., Zhang G., Cao G., *High-Density Surface EMG-Based Gesture Recognition Using a 3D Convolutional Neural Network*, Sensors, vol. 20, iss. 4, no. 1201 (2020), DOI: 10.3390/s20041201.

[26] Cote-Allard U., Nougarou F., Fall C.L., Giguere P., Gosselin C., Laviolette F., Gosselin B., *A convolutional neural network for robotic arm guidance using sEMG based frequency-features*, 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE Xplore, pp. 2464–2470 (2016), DOI: 10.1109/IROS.2016.7759384.

[27] Benalcazar M., Anchundia C.E., Zea J.A., Zambrano P., Jaramillo A.G., Segura M., *Real-Time Hand Gesture Recognition Based on Artificial Feed-Forward Neural Networks and EMG*, 2018 26th European Signal Processing Conference (EUSIPCO), IEEE Xplore, pp. 1492–1496 (2018), DOI: 10.23919/EUSIPCO.2018.8553126.

[28] Jaramillo-Yanez A., Benalcazar M.E., Mena-Maldonado E., *Real-Time Hand Gesture Recognition Using Surface Electromyography and Machine Learning: A Systematic Literature Review*, Sensors, vol. 20, iss. 9, no. 2467 (2020), DOI: 10.3390/s20092467.

[29] Zanghieri M., Benatti S., Burrello A., Kartsch V., Conti F., Benini L., *Robust Real-Time Embedded EMG Recognition Framework Using Temporal Convolutional Networks on a Multicore IoT Processor*, IEEE Transactions on Biomedical Circuits and Systems, vol. 14. no. 2, pp. 244–256 (2019), DOI: 10.1109/TBCAS.2019.2959160.

[30] Zhang Z., Yang K., Qian J., Zhang L., *Real-Time Surface EMG Pattern Recognition for Hand Gestures Based on an Artificial Neural Network*, Sensors, vol. 19, iss. 14, no. 3170 (2019), DOI: 10.3390/s19143170.

[31] Tavakoli M., Benussi C., Lopes P.A., Osorio L.B., de Almeida A.T., *Robust hand gesture recognition with a double channel surface EMG wearable armband and SVM classifier*, Biomedical Signal Processing and Control, vol. 46, pp. 121–130 (2018), DOI: 10.1016/j.bspc.2018.07.010.

[32] Fertl E., Castillo E., Stettinger G., Cuellar M.P., Morales D.P., *Hand Gesture Recognition on Edge Devices: Sensor Technologies, Algorithms, and Processing Hardware*, Sensors, vol. 25, iss. 6, no. 1687 (2025), DOI: 10.3390/s25061687.

[33] Nirmal S., Patil P., Kumar J.R.R., *CNN-AdaBoost based hybrid model for electricity theft detection in smart grid*, e-Prime-Advances in Electrical Engineering, Electronics and Energy, vol. 7, no. 100452 (2024), DOI: 10.1016/j.prime.2024.100452.

[34] Nelufule N., Masango M., Senamela P., Mawela H., Latakgomo M., Moloi P., *Digital forensics: A survey of emerging threats, challenges, and opportunities in smart grids*, 2024 IEEE 12th International Conference on Smart Energy Grid Engineering (SEGE), IEEE Xplore, pp. 109–114 (2024), DOI: 10.1109/SEGE62220.2024.10739495.

[35] Kawoosa A.I., Prashar D., Faheem M., Jha N., Khan A.A., *Using machine learning ensemble method for detection of energy theft in smart meters*, IET Generation, Transmission & Distribution, vol. 17, no. 21, pp. 4794–4809 (2023), DOI: 10.1049/gtd2.12997.

[36] Staszuk A., Wiatrak B., Tadeusiewicz R., Karuga-Kuźniewska E., Rybak Z., *Telerehabilitation approach for patients with hand impairment*, Acta of Bioengineering and Biomechanics, vol. 18, no. 4, pp. 55–62 (2016), DOI: 10.5277/ABB-00428-2015-03.

[37] Mohammad F., Al-Ahmadi S., Al-Muhtadi J., *RoGRUT: A Hybrid Deep Learning Model for Detecting Power Trapping in Smart Grids*, Computers, Materials & Continua, vol. 79, no. 2, pp. 3175–3192 (2024), DOI: 10.32604/cmc.2023.042873.

[38] Michałowska J., Tomilo P., *Measurement systems of electromagnetic field for aircraft with the use of a machine learning model*, Metrology and Measurement Systems, vol. 31, no. 3, pp. 577–594 (2024), DOI: 10.24425/mms.2024.150289.

[39] Calisan M., Olgun N., *Evaluation of classification performance of honey bee species with deep learning models*, Electrica, vol. 25, iss. 1, no. 0157 (2025), DOI: 10.5152/electr.2025.24157.

[40] Ji D., *Research on large-scale data anomaly detection based on edge computing and LSTM method*, International Journal of Low-Carbon Technologies, vol. 20, pp. 1292–1299 (2025), DOI: 10.1093/ijlct/ctaf086.

[41] Al Bataineh M., Abdoun D.I.A., Al Ahmad M., *Comparative Analysis of Machine Learning Algorithms for Antenna Alignments*, IEEE Access, vol. 13, pp. 114669–114680 (2025), DOI: 10.1109/ACCESS.2025.3585952.

[42] Lanka M.R., Liu G.R., *A novel neural network method for solid mechanics problems of heterogeneous bars*, Engineering Applications of Artificial Intelligence, vol. 156, iss. A, no. 111084 (2025), DOI: 10.1016/j.engappai.2025.111084.

[43] Huang L., Ma L., *Research on the Application of Intelligent Grading Method based on Improved ML Algorithm in Sustainable English Education*, Scalable Computing: Practice and Experience, vol. 25, no. 1, pp. 451–463 (2024), DOI: 10.12694/scpe.v25i1.2333.

[44] Geng W., Du Y., Jin W., Wei W., Hu Y., Li J., *Gesture recognition by instantaneous surface EMG images*, Scientific Reports, vol. 6, no. 36571 (2016), DOI: 10.1038/srep36571.

[45] Tepe C., Demir M.C., *Real-Time Classification of EMG Myo Armband Data Using Support Vector Machine*, IRBM, vol. 43, no. 4, pp. 300–308, (2022), DOI: 10.1016/j.irbm.2022.06.001.

[46] Farina D., Merletti R., Enoka R.M., *The extraction of neural strategies from the surface EMG: An update*, Journal of Applied Physiology, vol. 117, no. 11, pp. 1215–1230 (2014), DOI: 10.1152/japplphysiol.00162.2014.

[47] Atzori M., Gijsberts A., Castellini C., Caputo B., Hager A.G., Elsig S., Giatsidis G., Bassetto F., Muller H., *Electromyography data for non-invasive naturally-controlled robotic hand prostheses*, Scientific Data, vol. 1, no. 140053 (2014), DOI: 10.1038/sdata.2014.53.

[48] Hermens H.J., Freriks B., Disselhorst-Klug C., Rau G., *Development of recommendations for SEMG sensors and sensor placement procedures*, Journal of Electromyography and Kinesiology, vol. 10, no. 5, pp. 361–374 (2000), DOI: 10.1016/S1050-6411(00)00027-4.

[49] Liu X., Zhang Y., Wang Z., Chen J., Hu L., *Gesture Classification in Electromyography Signals for RealTime Control of Assistive Devices*, Bioengineering, vol. 10, iss. 11, no. 1324 (2023), DOI: 10.330/bioengineering10111324.

[50] Martinek R., Ladrova M., Sidikova M., Jaros R., Behbehani K., Kahankova R., Kawala-Sterniuk A., *Advanced Bioelectrical Signal Processing Methods: Past, Present, and Future Approach–Part III: Other Biosignals*, Sensors, vol. 21, iss. 18, no. 6064 (2021), DOI: 10.3390/s21186064.

[51] Zieliński T.P., *Cyfrowe przetwarzanie sygnałów od teorii do zastosowań*, Wydawnictwa Komunikacji i Łączności, Warszawa (2005).

[52] Irfan M., Ayub N., Althobiani F., Ali Z., Idrees M., Ullah S., Rahman S., Saeed Alwadie A., Ghonaim S.M., Abdushkour H., Alkahtani F.S., Alqhtani S., Gas P., *Energy theft identification using AdaBoost Ensembler in the Smart Grids*, CMC-Computers, Materials & Continua, vol. 72, no. 1, pp. 2141–2158 (2022), DOI: 10.32604/cmc.2022.025466.

[53] Chellamani G.K., Aishwarya N., Chandhana C., Kaur K., Babu R.T.S., *SpectroFusionNet a CNN approach utilizing spectrogram fusion for electric guitar play recognition*, Scientific Reports, vol. 15, no. 16842 (2025), DOI: 10.1038/s41598-025-00287-w.

[54] Siwek K., Swider S., *Prediction of photovoltaic energy generation using recurrent and transformer neural networks*, Archives of Electrical Engineering, vol. 74, no. 2, pp. 311–330 (2025), DOI: 10.24425/aee.2025.153900.

[55] Tran T.N., *Grid Search of Convolutional Neural Network model in the case of load forecasting*, Archives of Electrical Engineering, vol. 70, no. 1, pp. 25–36 (2021), DOI: 10.24425/aee.2021.136050.

[56] Gandolla M., Ferrante S., Ferrigno G., Baldassini D., Molteni F., Guanziroli E., Cotti Cottini M., Seneci C., Pedrocchi A., *Artificial neural network EMG classifier for functional hand grasp movements prediction*, Journal of International Medical Research, vol. 45, no. 6, pp. 1831–1847 (2016), DOI: 10.1177/0300060516656689.

[57] Naik G.R., Selvan S.E., Nguyen H.T., *Single-Channel EMG Classification with Ensemble-Empirical-Mode-Decomposition-Based ICA for Diagnosing Neuromuscular Disorders*, IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 24, no. 7, pp. 734–743 (2016), DOI: 10.1109/TNSRE.2015.2454503.

[58] Yaman E., Subasi A., *Comparison of Bagging and Boosting Ensemble Machine Learning Methods for Automated EMG Signal Classification*, BioMed Research International, vol. 2019, no. 9152506 (2019), DOI: 10.1155/2019/9152506.